A

Synopsis entitled

"Design And Development Of Machine Learning Algorithms For Soybean Leaf Disease Detection And Classification"

Submitted to

Swami Ramanand Teerth Marathwada University,

Vishnupuri, Nanded

in partial fulfilment for the Registration of the degree, Doctor of Philosophy in Computer Science (Faculty of Science & Technology)

Submitted By

Akshay Manik Shinde

School of Technology, S.R.T.M. University Nanded, Sub-Center, Peth, Latur

Under the guidance of

Dr. Santosh Prabhakar Shrikhande

Assistant Professor,

School of Technology,

Swami Ramanand Teerth Marathwada University, Sub-Centre, Peth, Latur, India.

December 2024

TABLE OF CONTENTS

Sr. No	Contents	Page No.
1	Abstract	3
2	Introduction	3-5
3	Literature Review	5-7
4	Significance of Study	7-8
5	Problem Statement	8-9
6	Research Objective	9
7	Research Methodology	10-11
8	Challenges and Limitations	11
9	Research Work Plan	12
9	References	12-13

1. Abstract:

Soybean crops are vital to global agriculture, and their productivity is significantly affected by diseases, which can lead to substantial yield losses. Timely and accurate disease detection is critical for effective crop management. This research work explores machine learning (ML) approaches for soybean leaf disease detection and classification. This work investigates various techniques such as supervised learning, deep learning, and feature extraction methods applied to images of soybean leaves. Several ML models, including support vector machines (SVM), convolutional neural networks (CNN), and random forests (RF), are compared to determine their efficiency in classifying different soybean diseases. Additionally, we examine the use of image pre-processing, data augmentation, and transfer learning to enhance model performance. The proposed models are evaluated on publicly available soybean leaf disease datasets, and performance metrics such as accuracy, precision, recall, and F1-score are used to assess their effectiveness. Results indicate that deep learning-based approaches, particularly CNNs, outperform traditional machine learning models in terms of classification accuracy. This research highlights the potential of ML techniques in automating disease detection, enabling early intervention and improving soybean crop management practices.

KEYWORDS: - Soybean Disease Detection, Soybean Disease Classification, Support Vector Machine [SVM], Machine Learning [ML]

2. Introduction

Soybean is one of the most widely cultivated legumes, as it can be processed into various foods and is an important global feed crop [1]. Disease is one of the main factors affecting soybean yields, with infected soybeans generally reducing yields by 10–30%, or more than 50% in severe cases. In particular, soybean leaf diseases can alter leaf colour and induce leaf loss which when severe can lower their disease resistance, thus decreasing crop yield and quality. Common and serious leaf diseases include soybean leaf spots and soybean rust. According to treatment plans, appropriate actions should be taken at the earliest possible stages of disease occurrence [2]. It is thus essential that we detect such diseases in a timely and accurate manner. Traditional disease identification requires manual intervention, and due to the complexity and diversity of leaf diseases, a great deal of experience is required to accurately identify the disease type, which makes it a time-consuming and labour-intensive process [3]. Numerous algorithms have been used to identify plant leaf diseases owing to the development of machine learning and

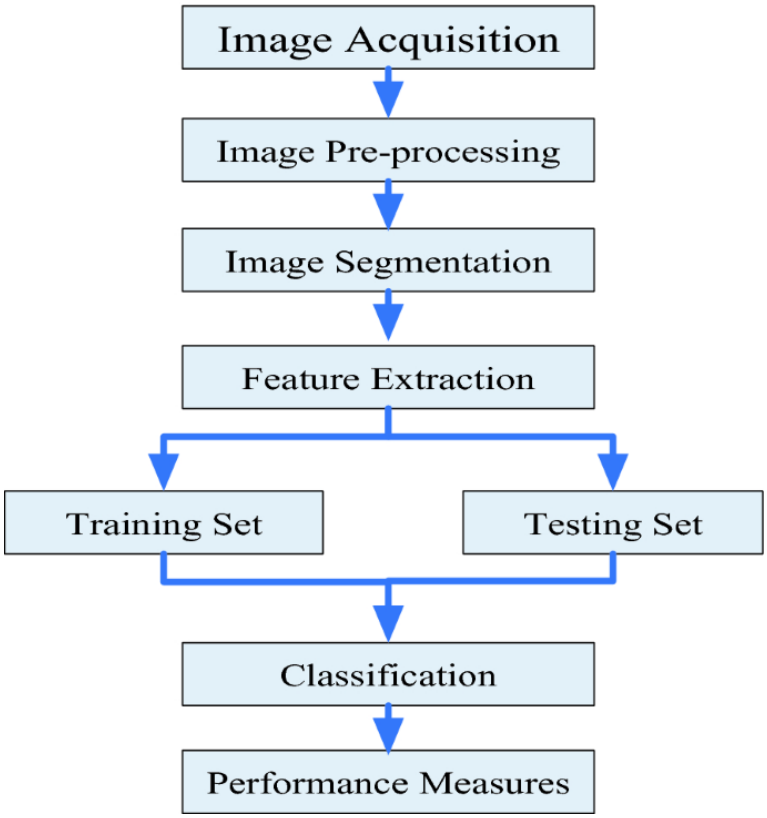
neural networks [4]. Currently, there are two primary approaches for plant leaf disease classification tasks: leaf disease recognition based on traditional machine learning algorithms and disease recognition based on deep learning.

Feature extraction and classification based on conventional machine learning algorithms is one of the traditional methods used to categorize plant leaf diseases. The benefits of this approach include a quick recognition speed and minimal hardware requirements.

To date, however, studies on soybean leaf disease recognition have been limited [11], with the majority focusing on the images of soybean leaf disease captured against a single background, and mostly using machine learning methods. In this study, soybean leaf images with complex background information were used as the training, validation, and testing samples, and an improved model-based soybean leaf disease image classification algorithm was proposed that addresses the issue of complex background interference typically present in real collections. The main findings of this study can be summarized as follows: (1) by performing image augmentation operations, such as random masking, on the dataset before training the network, the dataset was enriched to improve model robustness; (2) attention mechanisms were included at various depths to enable the ConvNeXt network to focus on discriminative features and reduce background interference; (3) the LeakyReLU activation function was used in the attention module to avoid situations in which the neurons did not learn when the input was negative; and (4) the improved network model was compared to existing models under the same conditions to validate its disease classification ability.

Following diagram shows process of Soybean Leaf Disease Detection and Classification

Soybean image Acquisition uses databases of Soybean Leaf Disease Dataset (PlantVillage), DeepSoil Dataset and PlantDoc dataset.



3. Literature Review:

A literature review of soybean leaf diseases focuses on the identification, diagnosis, and management of common diseases affecting soybean plants, as well as recent advancements in the use of technology for early detection and classification.

Panigrahi et al.[5] compared and analysed traditional machine learning algorithms such as naïve Bayes (NB), decision tree (DT), K-nearest neighbour (KNN), support vector machine (SVM), and random forest (RF) for soybean disease detection and after conduction experiment this research use plantvillage dataset after conduction of experiment Random Forest algorithm had the highest accuracy when compared to the other algorithms.

Rahaman et al.[6] used different machine learning algorithms such as Naïve Bays Classifier and k-nearest neighbour (KNN) classifier to detect and classify plant leaf diseases using texture features extracted from disease images, which was successful in detecting and identifying the selected diseases with had the highest accuracy 96.76% using k-nearest neighbour . The general steps of image classification using machine-learning algorithms are image preprocessing, feature extraction, and classifier training. Among these, feature extraction is the most crucial step as it directly impacts on the accuracy of the classification result.

Traditional feature extraction relies mainly on manual design, which can make feature extraction challenging [7].

Bansal et al.[10] combined three pre-trained models such as AlexNet , Google Net , Inception V3 to identify plant leaf and achieved satisfactory classification results.

Haque et al.[11] conducted experiments on common soybean diseases, using Google Net, Inception V3, and designing a new CNN network to classify and assess the severity of common soyabean diseases, respectively, and the proposed methods achieved good results in the identification of soyabean leaf diseases, and achieved certain results in the lightweighting of the model.

To date, however, studies on soybean leaf disease recognition have been limited, with the majority focusing on the images of soybean leaf disease captured against a single background, and mostly using machine learning methods [11]. In this study, soybean leaf images with complex background information were used as the training, validation, and testing samples, and an improved model-based soybean leaf disease image classification algorithm was proposed that addresses the issue of complex background interference typically present in real collections.

Soybean (Glycine max) is a crucial crop globally, providing high-protein food, oil, and animal feed. However, it is susceptible to various diseases that impact yield and quality. Efficient disease detection and classification are essential for precision agriculture, allowing farmers to take timely action to manage plant health. In recent years, advancements in image processing, machine learning, and deep learning have facilitated the development of systems for automated disease detection in soybean leaves. This literature review discusses the key methods, challenges, and findings in the domain of soybean leaf disease detection and classification.

Author Chirantan Chattopadhyay, S. J. Kolte, Farid Waliyar [9] described in his books Diseases of Edible Oilseed Crops presents an unprecedentedly thorough collection of information on the diseases of cultivated annual oilseed crops, including peanut, rapeseed-mustard, sesame, soybean, sunflower, and safflower describes different types of diseases on plant leaves.

The book focuses on major diseases of economic importance to each crop. Each chapter is devoted to a type of crop and a profile of affecting diseases according to geographical occurrence, epidemiology, symptoms, causal pathogens, host-pathogen interactions, biotechnological aspects.

Author Aluízio Borém, Felipe Lopes da Silva, Tuneo Sediyama, Willian Hytalo Ludke mentioned in his books to cluster in a single publication the most relevant and modern topics in soybean breeding. It is geared mainly to students and soybean breeders around the world. It is unique since it presents the challenges and opportunities faced by soybean breeders outside the temperate world [12].

4. SIGNIFICANCE OF THE STUDY

The significance of using Machine Learning (ML) algorithm for Soybean leaf disease detection and classification is to enhance the accuracy, efficiency, and scalability of detecting diseases affecting soybean crops. These ML techniques aim to automate the identification and classification of diseases from images of soybean leaves, facilitating early detection, effective management, and improved agricultural productivity. Specifically, the goals include

A] To Early Detection of Diseases:

- ML approaches enable the identification of diseases at an early stage, which is critical for preventing the spread of pathogens and minimizing crop damage.
- Timely detection helps farmers take preventive measures, such as applying pesticides or adjusting cultivation practices, to reduce the impact of diseases.

B] To Accurate Disease Classification:

- Different diseases affect soybean plants in distinct ways. ML models, especially those based on deep learning, can classify images of soybean leaves into categories, such as healthy or infected, and identify specific diseases.
- This classification improves the decision-making process in disease management by providing accurate diagnostics.

C] For Automation and Scalability:

- ML models can analyse large volumes of data (images of leaves) without the need for manual intervention, making it scalable for large-scale farms or automated monitoring systems.
- This reduces the time and labour required for disease detection, making it more feasible for farmers to monitor vast areas regularly.

D] To Improve Crop Management Process:

- By identifying and classifying diseases early, ML-driven systems can support farmers in making better-informed decisions on disease control, leading to higher yields and healthier crops.
- The model can also assist in integrating data from various sources (e.g., weather, soil condition) to offer comprehensive solutions.

E] For Reducing Reliance on Expert Knowledge:

- Machine learning reduces the dependency on expert agronomists or plant pathologists for disease identification. Farmers can use simple devices like smartphones or drones to capture leaf images, making the process more accessible and affordable.

F] To Precise Application of Treatments:

- Accurate detection enables the precise application of fungicides or other treatments, reducing waste and lowering the environmental impact of pesticides.

G] To Increased Yield and Profit:

- By using ML to detect diseases early and precisely, soybean growers can avoid significant losses, thereby improving the overall yield and profitability of the crop.

5. PROBLEM STATEMENT

The problem of soybean leaf diseases is a significant challenge for farmers, as various fungal, bacterial, and viral diseases can severely affect soybean plants, leading to reduced crop yields and quality. Common diseases such as Powdery Mildew, Rust, Anthracnose, and Bacterial Blight damage leaves, impair photosynthesis, and make plants more vulnerable to environmental stress. Early detection and accurate identification of these diseases are crucial for implementing effective treatments, reducing the need for excessive pesticide use, and minimizing crop loss. Timely intervention can help optimize soybean production and ensure food security.

Despite advancements in soybean leaf disease detection, there remains a gap in leveraging machine learning to enhance the accuracy, efficiency, and accessibility of soybean leaf disease detection. While machine learning models have shown promise in various agriculture applications, their potential for leaf disease prediction is underutilized due to challenges such

as data quality, model interpretability etc. Additionally, the variability in risk factors across various leaf diseases underscores the need for models that are not only accurate but also adaptable and inclusive.

The proposed research aims to address these gaps by designing and developing a robust, scalable, and user-friendly soybean leaf disease detection and classification system using machine learning algorithms. The system will focus on early detection through predictive modelling based on non-invasive or minimally invasive data, offering an accessible and efficient solution for agriculture material providers and farmers. By integrating advanced algorithms with explainable, the study seeks to improve both the technical performance and practical applicability of plant leaf disease detection and classification systems.

The problem at hand is to develop a Machine Learning (ML) algorithm that can automatically detect and classify soybean leaf diseases based on images of the leaves. This system should be able to identify multiple disease types and distinguish them from healthy leaves, enabling farmers to make informed decisions and optimize their crop management practices.

6. RESEARCH OBEJECTIVE

The objectives of the proposed research work are:

1) To study existing machine learning algorithms from the literature for Soybean leaf disease detection and classification

2) To design and develop the novel machine learning algorithm for soybean leaf disease detection and classification.

3) To apply the developed machine learning algorithm on soybean leaf disease datasets for measuring the detection and classification.

4) To compare the detection and classification accuracy of developed machine learning algorithm with existing machine learning algorithms

5) To improve the detection and classification accuracy of developed machine learning algorithm using different techniques or methodologies.

7. RESEARCH METHODOLOGY

The proposed work consists of the following components:

a. Literature Review:

- The proposed research work involves the study of various existing machine learning techniques/algorithms from literature for detection and classification of soybean leaf diseases.
- The research work involves the comparative study of various machine learning techniques from literature for detection and classification of soybean leaf diseases.

b. Data Set Used for Proposed Research Work:

- The proposed research work uses the Soybean Leaf Disease Dataset (PlantVillage): This dataset contains images of healthy and diseased soybean leaves with multiple disease classes is available on the popular platform i.e. Kaggle. This datset contains 61,486 images of plant leaves and backgrounds. The dataset includes images of 26 plant diseases.
- DeepSoil Dataset: Another prominent dataset that includes images of soybean plants under different environmental conditions, useful for both disease detection and plant health monitoring.
- PlantDoc Dataset: A large dataset of images of various plant diseases, including soybean. It contains 2569 images of leaves.

c. Research Design for Proposed Research Work:

- The proposed research work involves the designing and development of a novel methodology/technique using different machine learning algorithms for detection and classification of soybean leaf diseases.
- The new methodology will be designed using the guidelines of the standard research methodology and implemented in the **Python programming language** with different libraries available in the python.
- The developed methodology will be applied on different soybean leaf datasets and compare the accuracy of existing and new methodologies. The research will be carried out by using design and creation research strategy.

d.Software Tools Used for Proposed Research Work:

- Designing and developing a **Soybean leaf disease detection & classification** using machine learning algorithms involves a combination of various software tools. These tools are used for data collection, preprocessing, model building, evaluation, and deployment. Here's a list of some common software tools you might use:
- Programming Languages: The most widely used Python programming language will be used for conduction proposed research work, due to its vast libraries and frameworks (e.g., TensorFlow, Scikit-learn, Keras). The libraries used for proposed research work are Pandas, Scikit-learn, NumPy, XGBoost / LightGBM, Matplotlib / Seaborn etc. A robust library for implementing the basic machine learning algorithms like regression, classification.
- **Existing Algorithms Used:** There are several machine learning algorithms that are commonly used for **Soybean leaf disease detection & classification**. These algorithms help predict whether a leaf is likely to have disease based on various features. Below are some of the key algorithms used in leaf disese prediction:
    - **Decision Trees- Type:** Supervised, classification algorithm.
    - **Random Forest-Type:** Ensemble, classification algorithm.
    - **Support Vector Machine (SVM)-Type:** Supervised, classification algorithm.
    - **K-Nearest Neighbors (KNN)-Type:** Supervised, classification algorithm.
    - **Naive Bayes-Type:** Supervised, classification algorithm.
    - **K-Means Clustering-Type:** Unsupervised, clustering algorithm.

8. Challenges and Limitations:

- Data Imbalance: Some diseases may have fewer samples than others, making the model biased toward more frequent diseases.
- Environmental Factors: External factors such as lighting conditions and image quality can impact the model's performance.
- Generalization: The model may face challenges when generalized to different geographical regions with varying soybean strains and environmental conditions.

9. Research Work Plan: -

Phases	Plan of Work	Duration in Months					
		1-6	7-10	11-18	19-27	28-30	31-36
Phase I	Review of Literature	⟷	⟷	⟷	⟷	⟷	⟷
Phase II	Study of existing Methodology		⟷				
Phase III	Database creation/collection			⟷			
Phase IV	Design and Development of Machine Learning Model				⟷		
Phase V	Result and Conclusion					⟷	
Phase VI	Report writing						⟷

⟷ Arrow indicates the duration in months

10. References

1) Jianing, G. *et al.* Crispr/cas9 applications for improvement of soybeans, current scenarios, and future perspectives. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* **50**, 12678–12678 (2022).
2) Huang, W. *et al.* New optimized spectral indices for identifying and monitoring winter wheat diseases. *IEEE J. Sel. Top. Appl. Earth Obs Remote Sens.* **7**, 2516–2524 (2014).
3) Dhakal, A. & Shakya, S. Image-based plant disease detection with deep learning. *Int. J. Comput. Trends Technol.* **61**, 26–29 (2018).
4) Ngugi, L. C., Abelwahab, M. & Abo-Zahhad, M. Recent advances in image processing techniques for automated leaf pest and disease recognition-a review. *Inf. Process. Agric.* **8**, 27–51 (2021).
5) Hossain, E., Hossain, M. F. & Rahaman, M. A. A color and texture-based approach for the detection and classification of plant leaf disease using knn classifier. In *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 1–6 (IEEE, 2019).
6) Wagle, S. A. et al. Comparison of plant leaf classification using modified alexnet and support vector machine. Traitement du Signal **38** (2021).

7) Karlekar, A. & Seal, A. Soynet: Soybean leaf diseases classification. *Comput. Electron. Agric.* **172**, 105342 (2020).
8) Yu, M., Ma, X., Guan, H., Liu, M. & Zhang, T. A recognition method of soybean leaf diseases based on an improved deep learning model. *Front. Plant Sci.* **13**, 878834 (2022).
9) Diseases of Edible Oilseed Crops by Chirantan Chattopadhyay, S. J. Kolte, Farid Waliyar
10) Diseases of Field Crops and their Management by S. Parthasarathy, G. Thiribhuvanamala, K. Prabakar
11) Plant leaf disease prediction based on classification using machine learning techniques by Banu Priya M
12) Soybean Breeding by Felipe Lopes da Silva, Aluízio Borém, Tuneo Sediyama, Willian Hytalo Ludke

Akshay Manik Shinde
Research Scholar

Dr. Santosh P. Shrikhande
Research Supervisor / Guide